



## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **02082710 A**(43) Date of publication of application: **23.03.90**

(51) Int. Cl.

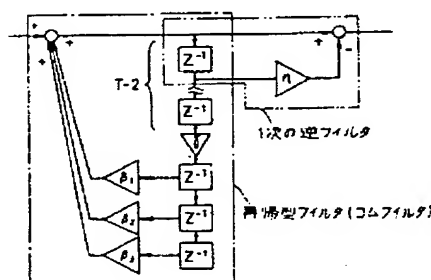
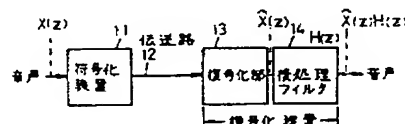
**H03H 17/02**  
**G10L 9/00**(21) Application number: **63235773**(22) Date of filing: **19.09.88**(71) Applicant: **NIPPON TELEGR & TELEPH  
CORP <NTT>**(72) Inventor: **MORIYA TAKEHIRO**(54) **AFTER-TREATMENT FILTER**

COPYRIGHT: (C)1990,JPO&amp;Japio

(57) Abstract:

PURPOSE: To improve the subjective quality of an encoding voice without damaging the naturalness and the clearness by combining a recurrence type filter to emphasize the part of a peak and a reverse filter to make blunt a smooth spectrum envelope with order lower than that of the recurrence type filter.

CONSTITUTION: A voice  $Xz$  is encoded by an encoder 11, a transmission line 12 is transmitted, the transmitted encoding information is decoded by a decoding part 13, a decoded output  $X'z$  is passed to an after-treatment filter 14 in which a characteristic is  $H_z$  and a voice output  $X'zH_z$  is obtained. The after-treatment filter 14 is composed of a return type filter to emphasize the part of the peak and the reverse filter to make blunt the low order smooth spectrum envelope from the degree of the recurrence type filter. Thus, without damaging the clearness and the naturalness, the subjective quality of the encoding voice can be improved.



⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A) 平2-82710

⑬ Int. Cl.<sup>3</sup>

H 03 H 17/02  
G 10 L 9/00

識別記号

庁内整理番号

N 8837-5 J  
H 8622-5 D

⑭ 公開 平成2年(1990)3月23日

審査請求 未請求 請求項の数 1 (全5頁)

⑮ 発明の名称 後処理フィルタ

⑯ 特 願 昭63-235773

⑰ 出 願 昭63(1988)9月19日

⑱ 発 明 者 守 谷 健 弘 東京都千代田区内幸町1丁目1番6号 日本電信電話株式会社内

⑲ 出 願 人 日本電信電話株式会社 東京都千代田区内幸町1丁目1番6号

⑳ 代 理 人 弁理士 草 野 卓

明 細 書

1. 発明の名称

後処理フィルタ

2. 特許請求の範囲

(1) 音声符号化復号化装置の復号化装置内で、再構成された音声信号に対してその短時間パワースペクトラムにおける声帯振動の基本周期に起因する山の成分と声道の共鳴に起因する山(ホルマント)の成分の両方またはどちらか一方を短時間ごとに適応的に強調する後処理フィルタにおいて、山の部分を強調する再帰型フィルタと、その再帰型フィルタの次数より低次の緩やかなスペクトル包絡をなまらせる逆フィルタとを組み合わせることを特徴とする後処理フィルタ。

3. 発明の詳細な説明

「産業上の利用分野」

この発明は音声波形系列を少ない情報量でより高い品質を保って符号化する方法のなかで、特に復号器側での後処理フィルタに関するものである。

「従来の技術」

高能率に情報圧縮された音声波形の符号化音声の量子化雑音を抑圧する方法として、種々の後処理フィルタが知られている。ピッチの調波構造を強調し、非周期的雑音を抑圧する方法としてはコームフィルタリング法(楕形フィルタ)が知られている。(長瀬裕実:コームフィルタリング雑音抑圧法の基本性能とPCM音声の品質改善への応用、電子通信学会論文誌、J65-A、No.6, pp. 572-578、1982)。この方法は周波数領域の処理で、周波数特性が櫛の歯の形になるようにしてピッチの調波構造を強調するものである。この方法では条件によっては処理により音声の周期性が強調されすぎて処理後の音声が反響的になったり、自然性や明瞭性が損なわれる場合があった。

一方ホルマント成分を強調するフィルタとしてはV. Ramamoorthy, et.al.: "Enhancement of ADPCM Speech Coding with Backward-Adaptive Algorithms for Postfiltering and Noise Feedback", IEEE, JSAC, vol. 6, pp. 364

— 382 (1988) でADPCM用後処理フィルタ(ポストフィルタ)が詳細に検討されている。特にこの文献の後半にはフィルタ特性が

$$G(z) = \frac{A^{(p)}(r_2 z)}{A^{(p)}(r_1 z)} \quad (1)$$

$A^{(p)}$  :  $p$  次の多項式

$r_1, r_2$  : 定数

$z$  : 遅延要素

の形式のフィルタで、ホルマントでない部分の量子化雑音の抑圧に効果をもつフィルタが提案されている。しかし  $r_1, r_2$  の差が小さいと効果が僅かになり、かと言って  $r_1$  と  $r_2$  との差が大きくなると異音が生じたり、自然性が損なわれる問題がある。

この発明の目的は情報圧縮して符号化された音声の主観的な品質を向上させるため、ピッチ周期の調波構造やホルマントの成分を強調し、かつ処理による副作用の殆どないような後処理をする後処理フィルタを提供することにある。

「課題を解決するための手段」

$$H(z) = \frac{(1 - \gamma z^{-1})}{(1 - \beta B(z))} \times \frac{A^{(m)}(r z)}{A^{(p)}(r z)} \quad (2)$$

$$A^{(p)}(r z) = \sum_{i=0}^p a_i^{(p)} (r z)^{-i} \quad (a_0 = 1)$$

$$B(z) = \beta_1 z^{-(T-1)} + \beta_2 z^{-T} + \beta_3 z^{-(T+1)}$$

ここで  $T, \beta_i, a_i$  はそれぞれ復号化された音声のピッチ周期、ピッチ予測係数、線形予測係数であり、 $r, \beta$  は実験的に適当に定める定数で、 $\gamma$  は  $\beta$  と  $B(z)$  に依存する係数である。

(2)式の第1項の分母は再帰型フィルタであるコムフィルタを構成し、ピッチ周期の調波構造強調で、大域的な傾きを分子の  $\gamma$  を係数とする1次のフィルタ(逆フィルタ)で補償する。つまり分母に3次のフィルタを採用することで、低域ではスペクトル性の山と谷の差が大きく、高域では差が小さくなるような自由度を持たせることができる。このため処理後の音声に反響的な歪が生じない。しかし3タップのフィルタは  $\beta$  の値に依存し

この発明によれば山の部分を強調する再帰型フィルタと、その再帰型フィルタの次数より低次の緩やかなスペクトル包絡をなまらせる逆フィルタとからなる。つまりこの発明によればピッチやホルマントの部分を再帰型フィルタにより局所的に強調すると同時にこれより低次の逆フィルタにより大域的なスペクトルの傾きが変化しないように適応的に補償する構造を持つことが特徴である。このため従来の後処理フィルタと異なり、処理による歪の付加が殆ど生じない。

「実施例」

この発明のフィルタは第1図に示すように音声符号化復号化装置の復号化器の中に組み込んで用いる。つまり音声  $X(z)$  は符号化装置11で符号化され、伝送路12を伝送され、その伝送された符号化情報は復号化部13で復号され、その復号された出力  $\hat{X}(z)$  は特性が  $H(z)$  の後処理フィルタ14に通され、音声出力  $\hat{X}(z)H(z)$  を得る。

この後処理フィルタ14の具体的な特性例を  $z$  変換を使って  $H(z)$  で表現する。

て大域的にスペクトルの傾きをもち、処理後の音声がかもった感じになる。これを防ぐために1次の傾きが0になるように、 $\beta$  と  $\delta$  にあわせて  $\gamma$  を決める。

第2項はホルマント強調フィルタになっており、次数の高い( $p$ 次)多項式からなる分母の再帰型フィルタでホルマントを強調し、次数の低い( $m$ 次)多項式からなる分子の逆フィルタで緩やかなスペクトル包絡をなまらせ、つまり傾きを補正する効果をもたせてある。両多項式ともに共通の短時間自己相関係数から求める。

これらの予測係数やピッチ周期は復号化された音声から求める場合と原音を分析して符号化され、伝送されたパラメータをそのまま用いる場合とがある。前者の場合には符号化方式に拘束されることがなく、柔軟な構成が可能であり、またパラメータを量子化する必要がない。しかしながら復号化装置内でのパラメータ推定演算が必要で、また量子化雑音の重畳した音声からパラメータを求める必要がある。後者の場合はこの逆で、復号器側で

の演算は不要で、原音から求めたパラメータを使うことができるが、パラメータには量子化歪や伝送路誤りが含まれる。また符号化器が予測係数やピッチ周期を補助情報として伝送する形式のものへの適用に限られる。

もちろん、例えばピッチのパラメータは前者、ホルマントのパラメータは後者の方法を用いる、あるいはその逆でもよい。また復号化した音声からパラメータを求める前者の場合、波形を再標本化してピッチ周期を高精度に求めることも可能である。H(z)のフィルタは稀れに処理後の音声の振幅(パワー)が処理前のものと著しく異なることがある。それを避けるためにフレーム単位(処理単位)ごとに処理後の音声の振幅(パワー)を処理前のものと合せる処理、つまり振幅(パワー)の正規化を併用することが望ましい。処理後の音声の滑らかにつながるように台形窓を使って処理後の音声を接続する。台形窓でのフレームを滑らかに接続する手法を併用するとフィルタによる副作用は殆ど感じられない。またこれらのパラメータは一

第5図は第4図と等価な処理を効率よく構成する方法を示す。左からの入力に対し、一番右の出力は $p$ 次の格子型再帰型フィルタの出力になっている。 $k_i$ は $i$ 次のPARCOR係数(特許754418号)である。すなわち一番右の出力はH(z)の第2項の分母のみの特性を示すものである。ところがH(z)の第2項のように $r$ を分子分母で共通に用い、かつPARCOR係数を用いると、 $p$ 次の合成フィルタの $m$ 次目の途中出力がそのままH(z)の第2項となる。予測係数 $\alpha$ のパラメータを直接用いる第4図の例では $\alpha$ の補間操作等により、フィルタが不安定になる場合があるが、第5図の例では $k$ が $\pm 1$ 以内で有れば必ず安定であり、パラメータの個数が $p$ 個で済むので実用上便利である。

第6図は第4図、第5図と等価なフィルタをLSPパラメータ(菅村、坂倉：線スペクトル対(LSP)音声分析合成方式による音声情報圧縮、電子通信学会論文誌J64-A, No. 8, pp. 599-606, (1981))を利用して実現したフィルタである。分母の多項式に対応するLSPパ

ラメータが $\theta$ であり、分子の多項式に対応するものが $\phi$ である。この構成もパラメータの補間による安定性はパラメータの大小関係で容易に判定できる。このほか様々な形式の線形フィルタで実現可能であることは明かである。

第2図は(1)原音、(2)後処理フィルタ前の符号化音声、(3)後処理フィルタの特性、及び(4)処理後の符号化音声の対数周波数特性(対数パワースペクトル)の例を示すものである。(2)で見られる調波構造の乱れが後処理フィルタを通した(4)で改善されていることが示されている。

第3図はフィルタのうち特にピッチ周期の調波構造を強調するH(z)の第1項の具体的構成例である。分母の特性を実現する3タップの再帰型フィルタ(コムフィルタ)と分子の特性を実現する1次の逆フィルタとにより構成される。

第4図はフィルタのうち特にホルマント強調を行うH(z)の第2項の具体的構成例である。これは分母の特性を実現する $p$ 次の再帰型フィルタと分子の特性を実現する $m$ 次の逆フィルタとを融合した形となる。

「発明の効果」

第7図は第1図の符号器復号器として重み付きベクトル量子化に基づく変換符号化(特願昭61-181770号)を用いてこの発明の主観的品質の改善効果を示すものである。(8KHz サンプル、8Kbit/s,  $r=0.95$ ,  $\delta=0.6$ )伝送路に1%のランダム誤りが生じる場合も試験に含めてある。縦軸は11種の入力音声に対する被験者28名の平均オビニオン値である。この結果、伝送路に誤りがある場合も無い場合にもこの発明によりPCM 0.5ビット相当の品質改善が実現されていることがわかる。

第8図は明瞭度試験(話者：3名、被験者：通話試験員4名)による日本語100音節の単音明瞭度の平均と標準偏差を示す。この結果より、こ

の発明の処理は明瞭度を劣化させることはない、つまり副作用がないことが分かる。すなわちこの発明の後処理フィルタは明瞭度や自然性を損なうことなく、符号化音声の主観的な品質を向上させる効果があることがわかる。

#### 4. 図面の簡単な説明

第1図はこの発明の後処理フィルタの音声符号化復号化装置に対する位置づけを表した図、第2図は(1)符号化前の音声、(2)符号化後の音声、(3)フィルタ  $H(z)$ 、(4)フィルタ通過後の音声の対数周波数特性図、第3図は  $H(z)$  の第1項の構成を表わした図、第4図は  $H(z)$  の第2項の構成を表わした図、第5図は  $H(z)$  の第2項をPARCOR係数を用いて構成した図、第6図は  $H(z)$  の第2項をLSPパラメータを用いて構成した図、第7図はこの発明の効果を対数PCM符号化と比較して平均オビニオン値で評価した図、第8図は明瞭度試験による単音明瞭度の平均値と標準偏差を表わした図である。

代理人 草野 卓

図1

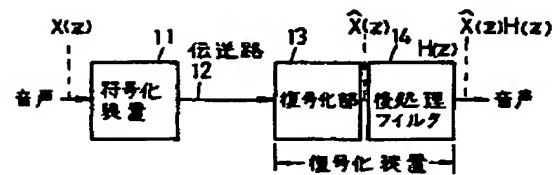


図2

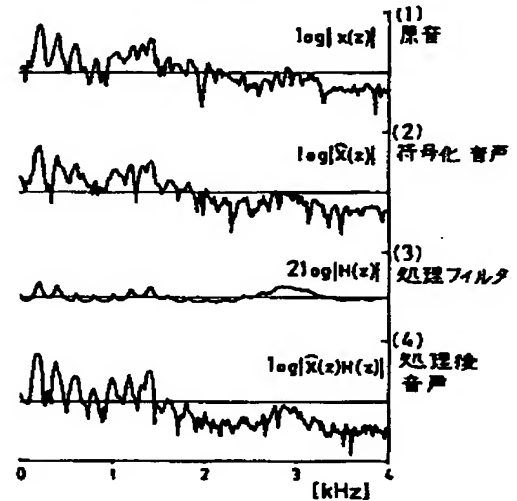


図3

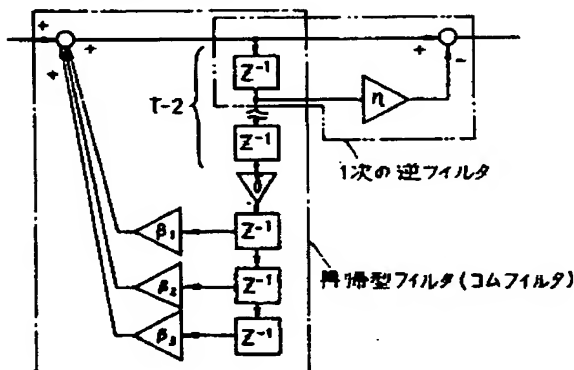


図4 P次の再帰型フィルタ

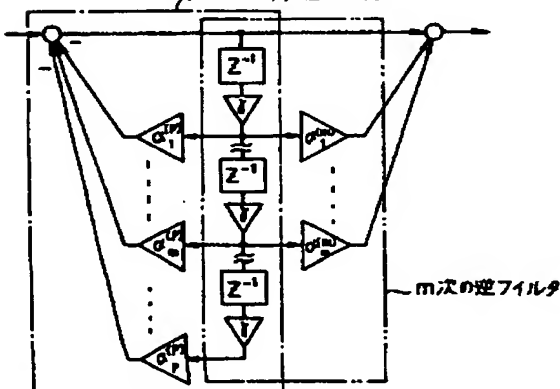


図7

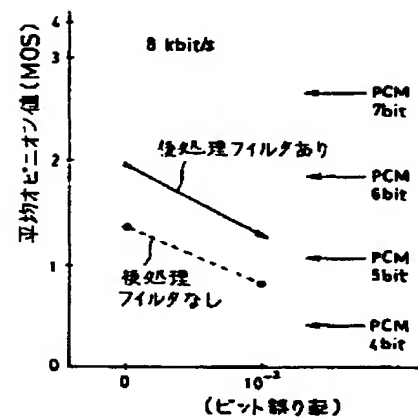
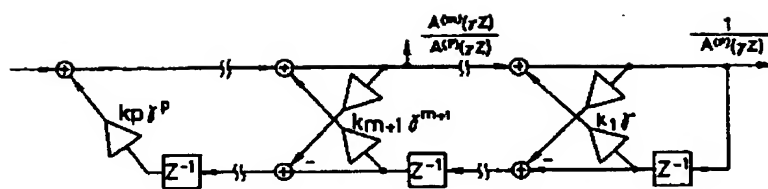


図8

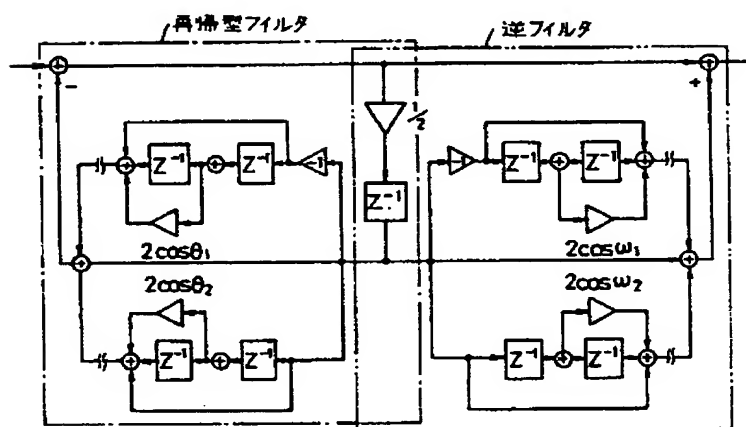
単 明 瞭 度 (括弧内は標準偏差)

後処理 フィルタ	なし	あり
誤り率 0%	93.1 (2.7)	94.1 (2.4)
誤り率 1%	92.5 (2.7)	92.9 (2.8)
原 音	97.6 (1.3)	98.1 (1.2)

カ 5 図



カ 6 図



A vocal tract is represented by  $p$  acoustics tubes having different inner diameters and equal lengths, which may be used as characteristic parameters of voice. A function using these parameters is called a cross-sectional area function for a vocal tract. Furthermore, it has been shown that PARCO coefficient  $k_i$  is equal to the reflection coefficient of the boundary surface of an acoustic tube. In this case, if the cross-sectional area of the  $i$ th acoustic tube from the lip is  $A_i$ , the relation between  $A_i$  and  $k_i$  is expressed by Eq.(2.8).

$$k_i = \frac{A_i - A_{i-1}}{A_i + A_{i-1}} \quad (2.8)$$

Assuming that the lip end is opened, that is,  $k_0 = -1$ , and the glottis is opened or closed, that is,  $k_{p+1} = 1$  or  $-1$ , as the boundary condition of a vocal tract acoustic tube, this vocal tract model is a lossless resonance system, and the acoustic tube has a line spectrum. A pair of line spectrum frequencies corresponding to this boundary condition is called a LSP<sup>7)</sup>.

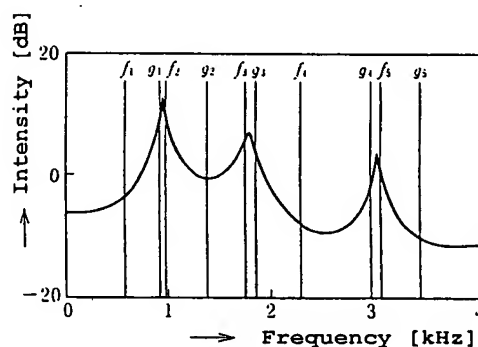


Fig. 2.7 Relations of Voice Spectrum and its Envelope to LSPs<sup>8)</sup>

The relations of a voice spectrum and its envelope to LSPs are shown in Fig.2.7<sup>8)</sup>. An order relation exists between the analyzed LSPs, and the LSP density represents the degree of concentration in frequency of the voice spectrum. Parts of the spectrum on which energy is concentrated correspond to the formants of the voice. 3 to 4 formants usually exist in voice and correspond to the resonance frequencies of a vocal tract. In order to derive formants from LSPs, an algorithm utilizing properties such as a frequency distribution of formants has been used.

LSPs represent well the characteristic of the envelope of a voice spectrum, and have been widely used as spectrum parameters in spectrum coding and hybrid coding methods which will be discussed in paragraph 2.4



国本音響学会 編

音響テクニカルシリーズ ①

# 音のコミュニケーション工学

—マルチメディア時代の音声・音・技術—

北脇 信彦 編著

コロナ社

$\phi(m)$  は式(2.3)で表される。

$$\phi(m) = \frac{1}{N} \sum_{n=1}^{N-|m|} x(m)w(n)x(n+|m|)w(n+|m|) \quad (2.3)$$

$$|m| = 0, 1, \dots, N-1$$

ここで、 $\{w(n)\}_{n=1}^N$  は、時間窓と呼ばれる音声波形切出し用の重み関数である。音声分析では、周波数分解能が高く、ひずみが少ないハミング窓が用いられることが多い。この自己相関関数  $\phi(m)$  から、音声スペクトル分析の多くのパラメータを導出することが可能であり、 $\phi(m)$  は音声分析の最も基本的なパラメータであるといえる。式(2.2)に対応する短時間パワースペクトル密度  $S(\omega)$  を  $\phi(m)$  を使って表すと、式(2.4)となる。

$$S(\omega) = \frac{1}{2\pi} \sum_{m=-(N-1)}^{N-1} \phi(m) \cos \omega m \quad (2.4)$$

式(2.1)に対応する離散系のフーリエ変換は、離散的フーリエ変換 (discrete Fourier transform, 略して DFT) と呼ばれる。 $S(\omega)$  は、音声の時系列信号  $x(n)$  の DFT によって求められるが、 $N$  が大きい場合には膨大な計算量が必要となる。そのため計算手順を工夫することによって計算量を飛躍的に減少させた計算法が、高速フーリエ変換 (fast Fourier transform, 略して FFT) と呼ばれるアルゴリズムである<sup>3)</sup>。

また、式(2.5)に示すように、短時間スペクトル  $S(\omega)$  の対数の逆フーリエ変換をとったものを、ケプストラム (cepstrum: spectrum の一部を反転した造語) という<sup>4)</sup>。

$$C(\tau) = \sum_{m=0}^M \log |S(\omega_m)|^2 \cos(\tau \omega_m) \quad (2.5)$$

$S(\omega)$  はスペクトル包絡成分と駆動音源成分の積であることから、式(2.5)のように対数をとることにより和の形に分解でき、ケプストラム分析によって、スペクトル包絡成分と駆動音源成分の両成分を分離して取り出すことができる。ケプストラム分析によるスペクトル包絡成分は、音声スペクトルの山と谷の情報をよく抽出しており、音声認識の特徴パラメータなどに用いられている。

### 2.3.3 線形予測分析

線形予測 (linear predictive coding, 略して LPC) 分析はパラメトリック分析の代表的な方法で、人間の音声生成モデルを設定し、このモデルを構成する特徴パラメータを推定することによってスペクトル分析を行う<sup>5)</sup>。すなわち、音声生成モデルとして、音声スペクトルは音源、声道伝達関数、口腔・鼻腔からの放射特性で表されると仮定する。

一般に、母音的な音の声道伝達関数はスペクトルの極で表され、それ以外の音は極と零によって表される。音声スペクトル  $T(\omega)$  を近似的に式(2.6)に示すような極だけで記述したものを全極モデルという。このとき、 $T(\omega)$  は全極形のスペクトル、すなわちスペクトルの山の特性を重視したスペクトルとなる。

$$\begin{aligned} T(\omega) &= \frac{1}{2\pi} |H(z)|^2 \\ &= \frac{1}{2\pi} \frac{\sigma^2}{\left|1 + \sum_{i=1}^p a_i z^{-i}\right|^2} \end{aligned} \quad (2.6)$$

ここで、 $\sigma$  は入力信号の実効値、 $p$  は極の数、 $a$  はスペクトル包絡パラメータである。 $a$  は、音声波形の標本値の自己相関関数を用いて、連立1次方程式を解くことによって求めることができる。

一方、音声波形  $x(n)$  が差分方程式(2.7)で生成されると仮定すれば、この生成モデルへの入力信号である  $\sigma e(n)$  から  $x(n)$  への伝達関数は全極形となり、式(2.6)と同一形となる。

$$x(n) + a_1 x(n-1) + \dots + a_p x(n-p) = \sigma e(n) \quad (2.7)$$

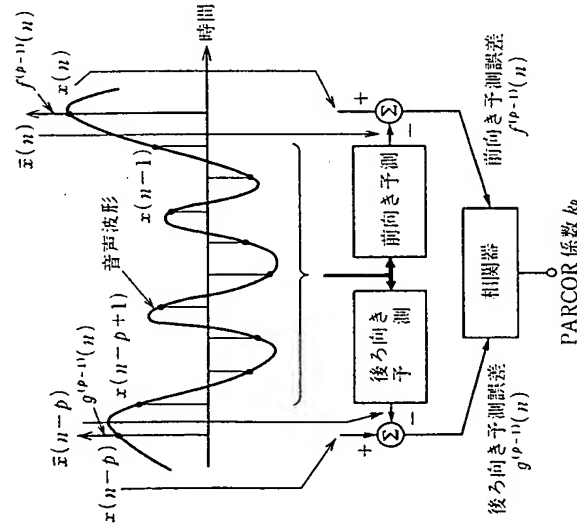
式(2.7)は、現時点の音声標本値  $x(n)$  が過去の  $p$  個の標本値  $x(n-1), x(n-2), \dots, x(n-p)$  の線形結合で予測されることを表しているともいえる。式(2.6)や式(2.7)の予測係数  $\{a_i\}_{i=1}^p$  と振幅  $\sigma$ 、あるいはそれと等価なパラメータを推定し、全極形スペクトルを決定する方法を線形予測分析という。これらのパラメータを決定するときの定式化の違いによって、最尤スペクトル推定法<sup>6)</sup>、共分散法、偏自己相関法<sup>7)</sup> (partial autocorrelation, 略して PAR

状は音声の特性に大きな影響を及ぼす。声道を  $p$  個の内径が異なる等しい長さの音響管で表し、これを音声の特徴パラメータとして用いることがある。これを声道の断面積関数という。また、PARCOR 係数  $k_i$  は、音響管の境界面における反射係数に等しいことが示されている。このとき、唇のほうから数えて  $i$  番目の音響管の断面積を  $A_i$  とすれば、 $A_i$  と  $k_i$  は式(2.8)の関係にある。

$$k_i = \frac{A_i - A_{i-1}}{A_i + A_{i-1}} \quad (2.8)$$

声道音響管の境界条件として、唇端を開放  $k_0 = -1$  とし、声門を開放  $k_{p+1} = 1$  および閉塞  $k_{p+1} = -1$  とすれば、この声道モデルは無損失の共振系となり、音響管は線スペクトルを持つ。この境界条件に対応する線スペクトル周波数の対を LSP という<sup>9)</sup>。

図 2.6 PARCOR 係数<sup>9)</sup>



COR), 線スペクトル対法<sup>9)</sup> (line spectrum pair, 略して LSP) などがある。

最尤スペクトル推定法や共分散法は、式(2.6)または式(2.7)における予測係数  $a$  と、振幅  $\sigma$  を直接推定する方法である。これに対して PARCOR 法は、図 2.6 に示すように、線形予測操作と相関操作の組合せにより、予測残差 (予測値と実測値との誤差) の相関係数の形でスペクトル包絡情報を抽出する<sup>9)</sup>。すなわち、音声波形の標本値  $x(n)$  と  $x(n-p)$  を、その間に挟まれた  $p-1$  個の標本値で前向きおよび後向き予測を行い、予測値  $\hat{x}(n)$  と  $\hat{x}(n-p)$  を得る<sup>9)</sup>。予測後の残差信号どうしの相関をとったものが PARCOR 係数である。

PARCOR 係数は、二つの標本値に挟まれた  $p-1$  個の標本値による線形の影響分を取り去った後の  $x(n)$  と  $x(n-p)$  の純粋な相関であるといえる。PARCOR 法によるスペクトル符号化方式は、合成器の安定性が保証されており、少ないビットレートで比較的良好な品質の音声を得られる。応用面では、階層を知らせるエレベータや、メッセージを伝える自動販売機などに実際に用いられた。

声道は、音声信号に音色 (スペクトル特性) を付与する調音器官で、その形

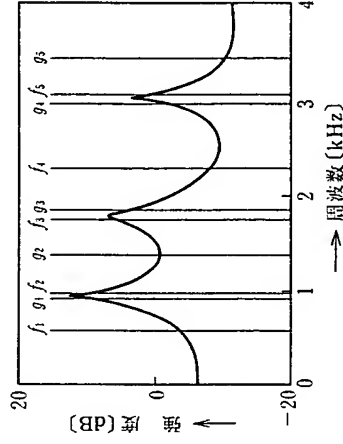


図 2.7 LSP と音声スペクトル、および包絡の関係<sup>9)</sup>

LSP と音声スペクトル、および包絡の関係を図 2.7 に示す<sup>9)</sup>。分析された LSP には順序関係が成立し、LSP の粗密により、音声スペクトルの周波数的な集中度が示される。スペクトル上でエネルギーの集中した部分は音声のホルマント (formant) に相当する。ホルマントは通常音声の中に 3~4 個存在し、声道の共振周波数と対応している。LSP からホルマントを導出するには、ホルマントの頻度分布などの性質を利用したアルゴリズムが用いられている。

LSP は、音声スペクトル包絡の特性をよく表しており、2.4 節で述べるスペクトル符号化やハイブリッド符号化方式におけるスペクトルパラメータとし

て広く用いられている。

### 2.3.4 ピッチ抽出

音声スペクトル分析は、音声の周波数スペクトルの構造のうち、声道の共振に対応するスペクトル包絡に関するものであった。もう一つの重要な特性は、音源に関する特性である。音源分析の中心は、声帯振動の有無と、声帯振動がある有声音の場合のピッチ周波数の抽出である。

声帯の振動には周期性があるが、必ずしも完全な周期性を持っているわけではない。また、駆動音源と声道とはたがいに独立ではないので、声道の影響を取り除いて声帯波を取り出すことはたいへん難しい。そのため、多くのピッチ抽出法が提案されている。大別すると、音声の波形処理によるもの、相関処理によるもの、スペクトル処理によるものがある。

音声のデジタル信号処理において最も広く用いられている方法が、相関処理による方法である。音声波形  $x(n)$  の自己相関関数  $\phi(m)$  は、ピッチ周期  $(T)$  と一致する点において顕著なピークを持つから、このピーク値を検出してピッチ抽出を行う。この方法が自己相関法によるピッチ抽出法である。

しかし、音声波形には駆動音源情報と声道情報が混在しているため、声道の影響を取り除いてからピッチ抽出を行ったほうが、抽出誤りが少なくなる。線形予測分析後の残差信号は、平たんなスペクトル包絡特性を持ち、駆動音源に基づくスペクトルの微細構造情報のみを持っている。したがって、残差信号の自己相関をとることによって、駆動音源の周期性を表すピークだけが明りょうに求められる。残差信号の自己相関関数からピッチ周波数を求める方法を、特に変形相関法と呼ぶことがある。

自己相関法と変形相関法によるピッチ周波数抽出の例を図 2.8 に示す<sup>2)</sup>。自己相関法では、声道の影響によるピークがたくさんあって、真のピッチ周波数を検出するのに誤りを生じる可能性があることがわかる。変形相関法では、声道の影響が取り除かれているため有声音の基本周期に基づくピークが顕著になっており、自己相関法よりもピッチ周波数の抽出誤りが少ない。

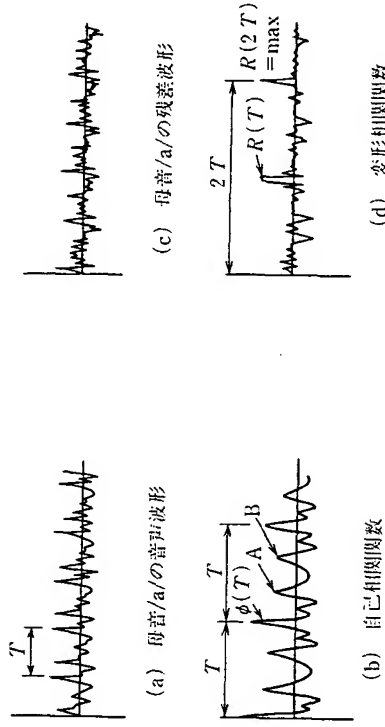


図 2.8 自己相関法と変形相関法によるピッチ周波数抽出の例<sup>2)</sup>

スペクトル処理によってピッチを抽出する代表的な方法が、式(2.5)で表したケプストラムを用いる方法である。この方法では、短時間パワースペクトル密度  $S(\omega)$  の対数をとることにより、スペクトル包絡成分と駆動音源成分を分離して、駆動音源成分からピッチ周波数を取り出している。

有声音と無声音の区別は、ピッチ周波数の分析に結び付けて行われる。すなわち、自己相関関数や変形相関関数は周期性の度合いを表しているから、このピーク値の大きさに基づいて有声音と無声音の判定が可能となる。

## 2.4 音声・楽音符号化の原理

### 2.4.1 低ビットレート符号化と品質

分析によって得られた特徴パラメータは、低ビットレートで量子化および伝送され、受信側でもとの信号に復元される。音声・楽音に対する符号化の品質とビットレートの関係を図 2.9 に示す<sup>3)</sup>。

音声・楽音の符号化には、大きく分けて音声波形を保存して伝送する方式 (speech waveform coding) と、音声スペクトルを保存して伝送する方式 (speech analysis-synthesis coding) とがある。音速通信で広く用いられている PCM や ADPCM (適応差分 PCM, adaptive differential PCM) は変形符

ATM	6, 246	FIR フィルタ	151	NOSFER	225
		[H]		[O]	
best-first 法	115	HMM	93	ODP 法	110
		HMM decomposition	109	OPINE	232
		[I]		OSS	183
CELP	30	IIR フィルタ	156	[P]	
CELP ハイブリッド符号化方式	33	IRS	225	PARCOR	25
COC	71	ITU-T	45	PCM	19, 31
CODEC	18	[K]		PSI-CELP	38
corrective training 法	103	KNN	104	PSQM	236
CS-ACELP	42	[L]		[R]	
CS-CELP	40	LD-CELP	36	RLS 法	162
CV	70	left-to-right	115	RPS	109
CVC	70	LMS アルゴリズム	161	[S]	
CYK アルゴリズム	118	LPC	25	SGD	109
		[D]		Stock Talk	131
DARPA	113	LR	225	[T]	
DCR	226	LR アルゴリズム	118	TTS	65
DFT	24	LR の客観測定器	231	TwinVQ	42
DP	93	LSP	26	[V]	
DTMF 信号	82	[M]		VCV	70
		MINT 法	199	VIP & P サービス実験	2
Early のアルゴリズム	118	MNRU	229	VQ 符号帳	99
ERLE	168	MPEG	45	VSELP	37
ES 射影法	168	MTF-STI	240		
		[N]			
FFT	24	NLMS アルゴリズム	161		
filtered-X アルゴリズム	197	NLMS 法	170		

△ ケプストラム 120

## — 著者略歴 —

北脇 信彦 (きたわき のぶひこ) 管 村 昇 (すがむら のぼる)  
 1971 年 東北大学大学院工学研究科修士課程修了 (電気および通信工学専攻)  
 1971 年 日本電信電話公社 (現 NTT) 勤務  
 1981 年 工学博士 (東北大学)  
 現在に至る

小泉 宣夫 (こいずみ のぶお)  
 1975 年 京都大学大学院工学研究科修士課程修了 (精密工学専攻)  
 1992 年 工学博士 (静岡大学)  
 1994 年 NTT データ通信 (株) 技術開発本部勤務  
 現在に至る

## 音のコミュニケーション工学

— マルチメディア時代の音声・音響技術 —

Speech Communication Technology

— Speech and Acoustics Technologies for Multimedia Services —

© (社) 日本音響学会 1996

1996 年 8 月 30 日 初版第 1 刷発行

## 検印省略

編 者 社団法人 日本音響学会  
 東京都渋谷区代々木 2-7-7  
 池田ビル 4 階  
 発行者 株式会社 コロナ社  
 代表者 牛来辰巳  
 印刷所 壮光舎印刷株式会社

112 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社  
 CORONA PUBLISHING CO., LTD.  
 Tokyo Japan

振替 00140-8-14844・電話 (03) 3941-3131 (代)

ISBN 4-339 01101-0

(製本: 染野製本所)

Printed in Japan



無断複写・転載を禁ずる

落丁・乱丁本はお取替えいたします

381/-